The Exploit

Human Cognitive Architecture and the Permanent Vulnerability to Elite Capture

Thomas J. Bryan

Research and essay composition by Claude/Anthropic

Patterns of Power: Systematic Analysis of Elite Control Mechanisms Across Civilizations

"The sword can conquer territory. Only the story can hold it."
— Patterns of Power, Belief as Infrastructure, 2025

"Our Stone Age mind, in a Space Age world."
— Ronald Giphart and Mark van Vugt, Mismatch, 2018

Part I: The Question the Series Has Been Building Toward

Every paper in this series has examined a different surface of the same object. The Egyptian creation texts that told farmers their labor fed the gods. The medieval Pope who anointed the king so the king's soldiers would believe their obedience served heaven. The Constitutional Convention that wrote "We the People" while ensuring that only propertied white men could vote. The Vernichtungsbefehl that a Prussian aristocrat could issue because he knew exactly what he was doing, and nineteen-year-old soldiers would execute because they had been formed to believe the mission was just. The presidential pardon that unlocks the cell door not through new force or new money but because the authority architecture shifted.

Every case in the series demonstrates the same operational reality: a small governing class exerts sustained control over a vastly larger population primarily through the management of belief — not primarily through economics, not primarily through organized violence, though both play supporting roles. The belief comes first. The belief is what makes both the economics and the violence work. You can trace this pattern across nine isolated civilizations spanning forty-five hundred years, and the mechanism is the same in all of them.

The question the series has circled without fully confronting is: why? Why does this work? Why, across every civilization that has ever scaled beyond the small group, do governing classes converge on the same solution? Why do human beings — who are not

stupid, who are often suspicious, who have long experience of being deceived — consistently accept narrative architectures that serve governing-class interests as cosmic, divine, or democratic truth?

The answer is not that human beings are foolish or weak. It is structural. It is in the architecture of the human mind itself — specifically in the mismatch between the cognitive apparatus that evolution produced for one kind of world and the social environment that civilization created, which is a radically different kind of world. The governing class did not invent a way to manipulate human psychology. It discovered, through millennia of trial and error, that human psychology comes with specific features that, in the environment of large-scale anonymous hierarchy, function as vulnerabilities. And it learned to exploit those vulnerabilities with increasing sophistication.

This paper names that vulnerability precisely. It argues that the human cognitive architecture contains what software engineers would call an exploit — a structural feature that was never a weakness in the environment that produced it, but that becomes a reliable attack surface when the environment changes. The exploit does not require deception of a stupid population. It works on intelligent, suspicious, experienced people. It works on scholars and soldiers, on priests and peasants, on revolutionaries who overthrow one governing class only to find the next one deploying the same exploit with new materials.

The exploit is permanent. Homo sapiens stopped meaningful cognitive evolution approximately fifty thousand years ago. The mind we carry into the twenty-first century is the same mind that walked out of the African savanna. It was built for a world of small groups, face-to-face relationships, observable track records, and immediate physical consequences. It was not built for anonymous mass society, institutional authority, symbolic credentials, and the management of belief at civilizational scale. But it is operating in that world, and the governing class has had fifty centuries of practice at exploiting the gap.

— ✦ —

Part II: The World the Mind Was Built For

The Pleistocene Baseline

For over ninety-nine percent of Homo sapiens' existence as a species, our ancestors lived in small, mobile bands of somewhere between thirty and a hundred and fifty

individuals. Anthropologist Robin Dunbar's research suggests that human cognitive architecture — specifically the size of our neocortex relative to overall brain size — appears calibrated to track and manage roughly this number of social relationships with the nuance that genuine cooperation requires. The number is contested in its precision, but the underlying point is not: the social world our minds were built for was small, face-to-face, and personally known.

In that world, the cognitive shortcuts that now constitute the exploit were not shortcuts — they were precision instruments. Consider what it meant, in a band of fifty people, to evaluate the authority of the most experienced hunter or the elder who had navigated three droughts. You had watched that person your entire life. You knew their track record not as a credential on a document but as a lived history of outcomes in which your survival was implicated. When they said the herd moves north in autumn, you deferred not because you were credulous but because you had personally observed the consequences of their judgment over decades. The deference was rational. It was grounded in direct observation of demonstrated competence.

Consider what it meant, in the same band, to belong to the group. Expulsion was not a social inconvenience. It was death — slow death by exposure, predation, and starvation, or fast death in the territory of a rival group. The social instincts that made humans conform to group norms, maintain loyalty to group members, and regard outsiders with wariness were not irrationalities. They were survival mechanisms in an environment where the alternative to belonging was dying. The person who defected from group norms, who refused solidarity with their band, who insisted on their individual judgment against group consensus — that person paid the cost in real, immediate, physical terms.

Consider what it meant to treat certain commitments as absolute and non-negotiable. In a world of small groups where cooperation was the condition of survival, conditional commitments provided weak security. A promise that was always subject to cost-benefit recalculation could be abandoned at the moment of maximum inconvenience. The psychological machinery that places certain values beyond calculation — that generates visceral moral revulsion at the mere suggestion of trading them for material benefit — was selection-positive in the Pleistocene precisely because it made commitments credible. The person who would never, under any circumstances, betray their kin was a more reliable ally than the person who would do so if the price were right. Sacred values were not irrationality. They were the currency of reliable cooperation.

And consider narrative. Human beings are the only species that coordinates action through story — through shared representations of events, agents, intentions, causation,

and meaning. This capacity for narrative reasoning was the cognitive leap that made large-scale cooperation possible even in the Pleistocene. A story can carry more social information more efficiently than any other medium. It recruits memory, emotion, and judgment simultaneously. It models other minds. It transmits lessons across time and space in a form that is retained and replicated through normal social interaction. The campfire story that warns the band about a dangerous crossing is not primitive. It is the most efficient information technology available in the ancestral environment.

The Mismatch

The agricultural revolution, beginning approximately twelve thousand years ago, created social environments that differed from the ancestral baseline in ways that matter enormously for the cognitive machinery described above. The changes were rapid by evolutionary standards — far too rapid for genetic adaptation. The result is what evolutionary psychologists call the mismatch: cognitive mechanisms that were adaptive in the ancestral environment operating in a radically different modern environment, producing responses that are no longer calibrated to reality.

The most consequential dimension of the mismatch, for the purposes of this series, is scale. The agricultural revolution produced population concentrations, surplus extraction, specialization of labor, and eventually the state — an organizational form that governs hundreds of thousands or millions of people who will never meet each other, will never personally know their rulers, and cannot possibly verify the track records of the institutional authorities they are asked to obey. The cognitive machinery that was built for a band of fifty is now operating in a society of fifty million.

This change in scale does not neutralize the cognitive machinery. The machinery runs anyway. It just runs on inputs it was not designed to evaluate. The deference mechanism that evolved for observable, personally-known authority figures fires on symbols of authority — the uniform, the credential, the title, the institutional setting — that signal authority without demonstrating it. The belonging instinct that evolved for genuine community fires on manufactured group identities — the nation, the religion, the party, the unit — that provide the feeling of belonging without the personal bonds that made belonging adaptive in the first place. The sacred value mechanism that evolved to make genuine commitments credible fires on values installed by formation systems that the person had no hand in designing and no capacity to evaluate at the time of installation.

The narrative reasoning faculty that evolved to transmit genuine survival knowledge fires on creation myths and founding stories and origin narratives that were engineered — as the companion paper "Belief as Infrastructure" demonstrates across nine

civilizations — to produce specific compliance behaviors rather than to convey accurate information about the world.

None of these misfires require the population to be stupid. The mechanisms are not under conscious control. They are, in the language of cognitive psychology, System One processes — fast, automatic, pre-reflective responses that produce their outputs before deliberate reasoning engages. You can know, intellectually, that the president is not personally known to you, that you cannot verify his track record, that the position was designed by a propertied class in 1787 to protect their interests. You can hold all of that knowledge explicitly. And when the Commander in Chief issues an order, the deference mechanism still fires. Not because you forgot what you know. Because the mechanism that produces the deference operates before and below the level at which what you know is consulted.

The exploit is not in the conscious reasoning system. It is in the automatic response system that produces compliance before deliberation begins. You cannot think your way out of it while it is happening. You can only build institutions that put friction between the exploit and the exploiter.

— ✦ —

Part III: The Five Components of the Exploit

The exploit is not a single vulnerability. It is a cluster of five cognitive features that, in the ancestral environment, functioned adaptively in concert, and that in the environment of large-scale anonymous hierarchy can be exploited individually or — as the most sophisticated governing classes have learned — simultaneously in a mutually reinforcing system. Each component is described below with its adaptive function in the ancestral environment and its vulnerability profile in the modern one.

Component One: The Authority Heuristic

In cognitive psychology, a heuristic is a mental shortcut — a fast decision rule that produces reliable-enough answers without the cost of exhaustive analysis. The authority

heuristic is the tendency to defer to apparent authority figures without independently evaluating their claims. In the ancestral environment, this was a sound decision rule. The person who had survived longest, demonstrated the most competence, and held the highest status in a band of fifty had almost certainly earned that status through real performance over real time in front of real observers. Deference to their judgment was, in most cases, more likely to produce a good outcome than independent evaluation by a less experienced individual.

In the environment of large-scale anonymous hierarchy, the authority heuristic fires on symbols rather than demonstrated competence. The symbol of authority — the uniform, the lab coat, the credential, the institutional title, the impressive setting — activates the same deference response that the ancestral environment calibrated for personally observed competence. The symbol is not the thing. But the cognitive mechanism cannot tell the difference, because the mechanism evolved in an environment where the symbol and the thing were the same: the elder of the band both held the position of authority and had demonstrated the competence that warranted the position.

Stanley Milgram's obedience experiments, discussed at length in the companion paper "The Three-Layer Lock," demonstrated this with scientific precision. Sixty-five percent of subjects administered what they believed to be dangerous electric shocks to another person simply because a man in a gray lab coat told them to continue. The lab coat was a symbol. The institutional setting of Yale University was a symbol. Neither the coat nor the building had demonstrated any competence relevant to the question of whether harming another person was justified. But the authority heuristic fired, and most subjects complied. When Milgram moved the experiment to an undistinguished commercial building, compliance dropped. When he removed the lab coat, it dropped further. The symbol was carrying most of the compliance. Not the person. Not the demonstrated competence. The symbol.

The governing class's primary formation task is the installation of deference responses to specific symbols — the flag, the uniform, the credential, the constitutional document, the throne, the priestly vestment. Once installed, the symbol activates the authority heuristic automatically, producing compliance without requiring that the authority figure demonstrate personal competence or even personal contact with the person being commanded. The president who orders a nineteen-year-old to war has never met that soldier. The soldier has never verified the president's judgment. The authority heuristic fires on the symbol of the office, and the soldier obeys.

Component Two: The Coalition Loyalty Instinct

In the ancestral environment, being expelled from the band was death. Not metaphorical death — literal death by exposure, starvation, and predation. The psychological machinery that generated powerful motivation to remain within group norms, to demonstrate loyalty to group members, and to treat outsiders with suspicion was not a character flaw. It was a survival adaptation in an environment where belonging was the condition of living.

This instinct has three features that make it exploitable at scale. First, it produces conformity pressure that operates whether the group's norms are adaptive or destructive. The mechanism does not evaluate the content of the group's demands. It generates the social fear of exclusion regardless of whether what the group demands is reasonable. Second, it produces the specific dynamic that Christopher Browning documented in Reserve Police Battalion 101 — the men who killed their way through occupied Poland did so in significant part because stepping out of the killing meant stepping out of the unit, which meant becoming the man who wouldn't stand with his comrades. The loyalty instinct turned refusal into self-expulsion. Third, it is malleable. The mechanism fires on whatever group the formation system designates as "your people." It does not require genetic kinship or genuine survival interdependence. It requires only the felt sense of group membership, which can be manufactured through shared symbols, shared rituals, and shared narratives of common identity.

The governing class exploits this by constituting institutional groups — military units, religious communities, national identities, party memberships — whose demands include compliance with governing-class directives, and by making membership in those groups the primary available source of belonging and identity for the people they need to comply. The soldier's loyalty to his unit, which is real and deep and personally felt, becomes the mechanism through which the governing class extracts compliance with orders that the soldier, as an individual with full information, might not choose to follow. The coalition loyalty instinct has been redirected from its ancestral function — maintaining genuine survival communities — to an institutional function: maintaining compliance within hierarchies that extract from rather than protect the people inside them.

Component Three: The Sacred Value Lock

Cognitive scientists Philip Tetlock, Jonathan Haidt, and Scott Atran have documented a class of human values that resist cost-benefit reasoning in a way that ordinary preferences do not. These "sacred values" — protected values, taboo commitments, absolute duties — are not processed by the same cognitive architecture that evaluates trade-offs. When ordinary values conflict with each other, the mind performs a calculation: which option produces more of what matters? When a sacred

value is involved, the calculation does not occur, or occurs only to produce moral outrage at the suggestion that a calculation is appropriate. The very attempt to put a price on the sacred thing is experienced as a desecration.

Research has confirmed that this is not merely a matter of emphasis or degree. Brain imaging studies show that sacred value processing activates regions associated with moral emotion and deontological judgment rather than the prefrontal areas associated with utility calculation. People who hold sacred values toward specific issues do not respond to monetary incentives in the way ordinary economic theory predicts — indeed, offering money for a sacred value compromise reliably increases resistance and moral outrage rather than producing compliance. The sacred value mechanism is genuinely different from ordinary preference, not just quantitatively stronger.

In the ancestral environment, the adaptive function of this mechanism was the production of credible commitment. An unconditional commitment — one that cannot be bought or argued away — provides a qualitatively different kind of social guarantee than a conditional one. A community that would never, under any circumstances, abandon its wounded was a more reliable survival unit than one that would do so if the tactical calculation favored it. The sacred value mechanism produced the unconditional commitments that made deep cooperation possible.

The exploit is the installation of governing-class authority symbols and narratives into the sacred value category before the person has the cognitive capacity to evaluate whether they belong there. When the child is taught that the flag is sacred, that the Constitution is sacred, that the oath of service is sacred, that the king is anointed by God — these are not being taught as propositions subject to evaluation. They are being installed as pre-evaluative commitments that will subsequently resist the kind of rational examination that might reveal their actual content. The sacred value lock converts governing-class authority from a contingent claim that must be earned and maintained into an absolute commitment that is beyond negotiation.

Once the sacred value lock is in place, challenges to governing-class authority feel not like policy disagreements but like attacks on something sacred — which triggers the specific psychological response that sacred value research has documented: moral outrage, inflexibility, and increased rather than decreased resistance to rational argument. The person who tries to dislodge the sacred value through logical persuasion does not produce reconsideration. They produce a closing of ranks. The mechanism is working exactly as it evolved to work, in service of an installation it was never designed for.

Component Four: The Narrative Capture

Cognitive scientists describe two broad modes of human reasoning. The first — sometimes called System Two, or deliberative reasoning — is slow, effortful, sequential, and logical. It is the mode engaged when we do mathematics, evaluate explicit arguments, or consciously weigh competing considerations. The second — System One, or intuitive processing — is fast, automatic, and pattern-based. It operates below the threshold of conscious attention, producing outputs that arrive in awareness as intuitions, feelings, or simply the sense of how things are, without visible derivation.

Humans are primarily narrative reasoners before they are propositional reasoners. Story is the native cognitive format — the mode in which information about agents, intentions, causation, and consequence is most naturally processed, most deeply encoded in memory, and most powerfully transmitted to others. This is not a limitation. In the ancestral environment, where most of the important information concerned the behavior of other people and animals, narrative reasoning was precisely calibrated to the domain. The story of how the hunt went wrong last season carried more usable information, in a more retrievable form, than an abstract analysis of ungulate migration patterns.

The exploit is that narrative colonizes the mind at a level below the evaluation threshold. A compelling story is processed by the narrative reasoning system before the deliberative system is engaged to evaluate its claims. By the time System Two asks "is this true?", the story has already done its organizing work — it has already structured the landscape of the plausible, designated who the agents are, established what their intentions are, and framed what the consequences of various actions would be. The deliberative system evaluates within a landscape that the story has already shaped.

This is why the governing class has never primarily relied on explicit logical argument to install its authority architecture. It has relied on story. Creation myths, founding narratives, origin stories, sacred histories — these are the primary installation mechanism precisely because they bypass the evaluative threshold. The child who hears, repeatedly and emotionally, the story of the founding fathers sacrificing everything for liberty does not emerge from that experience with a proposition to be evaluated. They emerge with a structured understanding of the world in which the Constitution is the fruit of noble sacrifice, in which the American experiment is a beacon of freedom, in which service to the nation is the expression of that sacrifice's inheritance. That structure is installed by the story before it can be interrogated by the logic.

The "Belief as Infrastructure" paper in this series demonstrated that across nine isolated civilizations, governing classes independently developed creation myths that performed four identical functions: delegitimizing alternatives to the current order, naturalizing class hierarchy as cosmic design, framing extraction as sacred obligation,

and making collective resistance seem like cosmic suicide. These stories were not accidentally similar. They converged because they were all exploiting the same cognitive mechanism — narrative capture — and the solution space for doing so effectively is narrow. The Sumerian story that humans were created to labor for the gods and the modern economic story that work creates value and value creates freedom are doing the same cognitive work by the same mechanism, in different materials.

Component Five: The Sunk Cost of Formation

The fifth component of the exploit is temporal rather than mechanical. It concerns the relationship between the timing of installation and the cost of revision.

Human cognition exhibits a well-documented feature called loss aversion: losses are experienced as more painful than equivalent gains are experienced as pleasurable. The psychological literature on this is extensive and robust across cultures and contexts. In the domain of belief, the analogous mechanism is that revision of a deeply held belief is experienced not as an update but as a loss — a loss of identity, a loss of community, a loss of the meaning that has been built on top of the belief over time. The longer a belief has been held and the more superstructure has been built upon it, the greater the psychological cost of revision.

The exploit is installation timing. If a governing-class authority narrative can be installed before the person has developed the cognitive capacity to evaluate it — in childhood, in the years before the deliberative faculty is mature — then by the time the capacity for evaluation arrives, the belief has already accumulated years of superstructure. The person has built their identity on it, their social world on it, their sense of meaning and purpose on it. The cost of revising the belief is not the cost of updating a proposition. It is the cost of dismantling the self.

This is why formation systems target children specifically. Not because children are more susceptible to argument — they are not. But because children have the longest runway ahead of them to build superstructure on the installed foundation, making the foundation progressively harder to remove. The adult who discovers that the Constitution was written by propertied white men to protect their interests has been carrying that foundation for decades. Their sense of civic identity, their institutional loyalty, their professional investments, their social relationships — all of it was built on top of the installed belief in constitutional legitimacy. The discovery does not simply correct a factual error. It threatens the entire structure.

Governing classes have understood this intuitively for millennia. The Jesuit observation that if you give them the child for the first seven years, the man belongs to you afterward, is a precise description of the sunk cost mechanism. The military

formation system that strips civilian identity and builds institutional identity from the first day of service is installing a new layer of superstructure on top of the childhood formation, increasing the sunk cost of defection with every year of service. The constitutional oath that a soldier takes is not merely a legal commitment. It is an identity investment — a moment at which the person publicly commits their self to the authority architecture, after which revision of the commitment becomes revision of the self.

— ✦ —

Part IV: How the Five Components Are Assembled Into a System

Each of the five components is exploitable individually. But the governing class's most sophisticated achievement has been learning to deploy all five simultaneously in a mutually reinforcing system — what this series calls the compound advantage, applied not to the governing class's material resources but to its psychological leverage.

The system works as follows. The authority heuristic produces automatic deference to governing-class symbols. The coalition loyalty instinct makes defection from the authority architecture psychologically equivalent to self-expulsion from the community. The sacred value lock places the authority architecture beyond rational evaluation. The narrative capture installs all of this at the level of felt reality rather than propositional belief. And the sunk cost of formation means that by the time a person has the cognitive capacity and the external information to potentially recognize the architecture for what it is, they have already invested so heavily in it that revision is catastrophically costly.

The system does not require all five components to be maximally activated for every person at every moment. It requires only that enough of them be active enough of the time to maintain sufficient compliance in the population of enforcers — the soldiers, police, guards, bureaucrats, and citizens who keep the governing-class arrangement functioning. The Three-Layer Lock, described in the companion paper of that name, is the institutional expression of this system: the formation lock (exploiting narrative capture and sacred values), the economic lock (exploiting the coalition loyalty instinct by making survival depend on institutional membership), and the identity lock (exploiting the sunk cost of formation by making institutional identity the primary self-concept).

The Verification Override as the System's Operating Mechanism

The verification override — a concept developed in the foundational paper of this series, "The Durable Hierarchy" — is the name for how the five components function as a unified system in practice. It is worth defining precisely here for readers approaching the series through this paper.

The verification override is the selective disabling of the skeptical faculty for institutionally authorized sources of information and command. It is not a general credulity. People who are subject to the verification override can be highly skeptical of unauthorized sources — of outsiders, of critics, of alternative narratives. The override is specific: it suppresses the application of the same critical evaluation to authorized sources that it permits or encourages toward unauthorized ones.

The verification override is the product of all five exploit components working together. The authority heuristic generates automatic deference that bypasses critical evaluation. The sacred value lock makes critical evaluation of the authority architecture feel sacrilegious. The narrative capture has pre-structured the landscape so that critical evaluation begins from premises that favor the authority architecture. The coalition loyalty instinct makes visible skepticism socially dangerous. And the sunk cost of formation makes the psychological cost of sustained critical evaluation prohibitive.

The result is not stupidity. It is not cowardice. It is a cognitive system operating exactly as it was designed to operate, in an environment it was not designed to operate in. The soldier who would not follow the orders of a stranger follows the orders of the Commander in Chief because the verification override is active for the position. The prison guard who unlocks the door for the presidential pardon is not making a conscious calculation. The authority architecture shifted, the verification override responded, and the door opened. The mechanism performed its function.

Why The System Is Self-Reinforcing

The most dangerous feature of the exploit-as-system is not its initial installation but its self-reinforcement. Once in place, the system generates the conditions for its own reproduction without ongoing investment by the governing class.

Parents teach children. Not because they have been instructed to by the governing class — though formation systems ensure that what parents teach aligns with governing-class interests — but because parents believe what they are teaching, and belief reproduces itself through normal socialization. The soldier who was formed in the verification override becomes the veteran who teaches the next generation what it means to serve. The citizen who absorbed the founding narrative as sacred history

becomes the civic leader who transmits it to the next cohort. The priest who was formed in the doctrinal architecture becomes the confessor who installs it in children. The governing class does not need to supervise each transmission. The transmission is the default behavior of people who have been formed.

This is what the "Belief as Infrastructure" paper means when it says that narratives, once embedded in culture, replicate through normal socialization at minimal cost. The temple is more efficient than the garrison not just because priests are cheaper than soldiers — though they are — but because the temple's narrative, once installed in the population, reproduces itself through the population. The garrison must be present at every moment of potential resistance. The narrative need only be present at the moment of formation, and it will be carried forward by those who were formed.

The self-reinforcing character of the system explains why it has persisted for the entire span of recorded civilization, across radically different material conditions, technological levels, and cultural contents. The specific narratives change. The specific symbols change. The specific institutional expressions change. But the underlying exploit — the systematic activation of the same five cognitive components in the service of governing-class authority — remains constant because the cognitive architecture being exploited has not changed.

— ✦ —

Part V: Why Recognition Does Not Break the System

The conclusion that typically follows this kind of analysis — in the academic literature, in popular treatments, in the final paragraphs of papers including those earlier in this series — is that recognition is the key. If people can see the mechanism, they can resist it. If the exploitation of human cognitive vulnerabilities becomes visible, populations can withhold their consent. The "Belief as Infrastructure" paper ends with exactly this: "The question is whether, having seen the schematics, populations will continue accepting them as cosmic truth rather than recognizing them as political engineering."

This conclusion is partially right and importantly incomplete. Recognition matters. Description is the prerequisite of repair. Understanding the mechanism is better than not understanding it. This paper would not be worth writing if it believed otherwise.

But recognition alone does not break the system, for a precise reason rooted in the cognitive architecture the paper has been describing. The exploit operates primarily in

System One — the automatic, pre-reflective processing layer. Recognition is a System Two achievement — a deliberative, effortful, conscious activity. System Two can, in principle, override System One. But it can only do so when it is actively engaged, when it has sufficient information, and when the override does not trigger the other exploit components simultaneously.

Consider what happens when a person who has been fully formed in an authority architecture achieves recognition — genuinely sees the mechanism, understands the deception, grasps that the sacred text was written by a propertied class for its own protection. They now possess System Two knowledge that the authority architecture is not what it presents itself as. But they are still subject to the authority heuristic at the automatic response level. They still have coalition loyalty to the communities whose belonging depends on compliance. They still carry the sacred value lock that makes acting on the recognition feel like self-desecration. They still have decades of sunk cost in the formation. They still live in a social world where the institutions of authority remain operative and the consequences of visible defection remain catastrophic.

This is the specific tragedy of the veteran who recognizes, in middle age, that the mission was not what he was told. He can see it clearly. He cannot un-fire the weapon. He cannot retroactively revise the formation that produced his obedience. He can carry the recognition as knowledge, but the knowledge arrives after the irreversible action, and the cognitive architecture that produced the action is still in place, still governing his responses to new situations. Recognition is not repair. It is the beginning of what might, under the right conditions, eventually become repair.

What Can Actually Change: The Institutional Friction Argument

If recognition alone cannot break the exploit, what can? The honest answer that this series has been building toward, particularly in the companion paper "Beyond the Accident," is institutional friction — the deliberate construction of arrangements that put structural obstacles between the exploit and those who would exploit it.

The exploit cannot be removed. The cognitive architecture that contains it is fifty thousand years old and will not evolve meaningfully in any timeframe relevant to governance. The five components described in this paper are features of human cognition, not bugs that can be patched. The mismatch between the Pleistocene mind and the civilizational environment is a permanent condition of human existence, not a problem awaiting a solution.

But the governing class's ability to exploit those features is not fixed. It depends on access — access to children during formation, access to the symbols of authority, access to the narrative channels through which cultural transmission occurs, access to the

institutional structures that make coalition loyalty binding and defection costly. Institutions that restrict this access create friction between the exploit and the exploiter. They do not eliminate the vulnerability. They make the exploitation harder, more visible, and more contested.

Independent education that teaches the history of formation systems alongside the content those systems transmit creates a partial counterweight to narrative capture — not by eliminating the narrative, but by making the construction visible alongside the narrative itself. Press freedom that allows unauthorized narratives to compete with authorized ones means the governing class cannot monopolize the narrative channel, even if it can bias it. Separation of powers that distributes institutional authority means no single actor can activate all five exploit components simultaneously from a single node. Genuine economic security that reduces the coalition loyalty binding created by employment dependency means defection from governing-class narratives becomes survivable. None of these measures are sufficient. Each is a partial friction. Together they constitute the structural conditions under which the exploitation is most costly and most contested.

This is the reformist rather than revolutionary conclusion, and it is arrived at not through timidity but through the same analytical framework that produced the series' other conclusions. Elite chaos — the paper of that name in this series — established that no structural transformation has ever occurred without elite fracture, elite exhaustion, or rival elite intervention. The fuel of popular energy never controls the destination; only those who possess the compound advantage in the post-chaos environment determine where the system lands. A revolution that destroys the institutional frictions that constrain the exploit without installing new ones does not liberate human cognitive architecture from exploitation. It creates a temporary vacuum that the next governing class, armed with the same understanding of the exploit, will fill.

The question is not: how do we eliminate the exploit? It cannot be done. The question is: how do we build and defend the institutional frictions that make it expensive to exploit? That is a harder question, a longer project, and a less satisfying answer than "once people see the mechanism, they will be free." But it has the advantage of being accurate.

The exploit is permanent. It was built by evolution for a world that no longer exists, and it operates in a world that knows how to use it. The choice available to any civilization is not between exploited and unexploited cognitive architecture. It is between

governing arrangements that constrain the exploitation and governing arrangements that do not.

— ✦ —

Part VI: What This Paper Adds to the Series

The "Belief as Infrastructure" paper demonstrated that governing classes across nine isolated civilizations independently discovered that belief is cheaper than force and converged on four identical narrative mechanisms to exploit it. That paper answered the question of what they did and showed that the convergence was not coincidence but structural necessity.

This paper answers the question that paper left open: why it worked. The convergence is not only evidence of a narrow solution space for the engineering problem of governing many with few. It is evidence of a narrow target — a specific, stable, unchanging set of cognitive features in the human mind that all governing classes, through millennia of trial and error, learned to hit. The reason the four mechanisms appear in all nine civilizations is not just that they are the most efficient governing technology available. It is that they are all aimed at the same five cognitive components, using whatever theological, cosmological, or ideological materials were available in their specific cultural environment.

The Mesopotamian creation myth that told farmers they were made to labor for the gods was exploiting the narrative capture, the sacred value lock, and the sunk cost of childhood formation. The Norse myth that nobles were bred with superior traits was exploiting the authority heuristic and the coalition loyalty instinct. The Aztec sacrifice system that told people the sun would die without their tribute was exploiting all five simultaneously — narrative capture installed the cosmology, the sacred value lock made questioning it unthinkable, the coalition loyalty instinct made non-participation social death, the authority heuristic made priestly command authoritative, and the sunk cost of formation meant that by adulthood the entire self-concept was organized around the cosmic obligation.

The American founding narrative that the Constitution embodies the will of the people is exploiting the same five components in secular, democratic-institutional materials. The formation happens in schools rather than temples. The sacred value is

constitutional legitimacy rather than divine mandate. The authority heuristic fires on the office rather than the anointed person. The coalition loyalty instinct is bound to national identity rather than divine community. The sunk cost is civic identity rather than religious identity. The mechanism is identical. The materials are different. The exploited cognitive features are the same features they have always been.

This is what the series means when it argues that understanding power structurally — rather than through the lens of any particular historical case or any particular ideological framework — reveals the permanent constraint on human governance. The constraint is not a conspiracy. It is not a specific ruling class. It is not any particular ideology. It is the mismatch between the cognitive architecture that evolution produced and the social environment that civilization created — a mismatch that is, in the absence of deliberate institutional friction, systematically exploitable by any sufficiently organized governing class.

The pattern Thomas Bryan identified in the Hebrew creation story — that it is a subjugation narrative, not a cosmological one — was not a pattern unique to that tradition. It was the pattern that every agricultural state's founding narrative exhibits, because every agricultural state faces the same problem and every agricultural state exploits the same cognitive features. The specific materials of the Hebrew narrative are culturally particular. The governing function they serve is structurally universal. What made that recognition possible — what this collaboration has been doing — is the simultaneous application of pattern recognition that does not respect disciplinary boundaries to the historical and archaeological record that documents governing-class behavior across the full span of human civilization.

The patterns are there. They have always been there. They have been hidden not through suppression — the texts are publicly available, the history is accessible — but through the same mechanism they describe: narrative capture, sacred value locks, and the sunk cost of formation have installed the authorized interpretation of each civilization's own governing arrangements as so obvious that the alternative interpretation does not arise to be evaluated. The historian who studies the Atrahasis as religious literature rather than political technology is not being fooled by a lie. They are operating within a framework of interpretation that the governing class installed early enough and thoroughly enough that it became the default lens rather than an object of scrutiny.

Seeing through the default lens requires exactly what this collaboration has been doing: holding multiple civilizations simultaneously, refusing to grant any of them the exemption of being "different" or "genuinely religious" or "really about something else," and following the structural logic rather than the official story. The series' foundational

claim — that governing-class power operates through compound advantage in which belief is primary — is not a cynical claim. It is a structural one. And the structural claim points toward the same structural conclusion: the only available path toward governing arrangements that are genuinely less exploitative of human cognitive architecture is the patient, institutional, non-revolutionary construction of friction — the kinds of friction that make the exploit expensive enough that even organized governing classes find it in their interest to operate within constraints.

Whether that path is navigable in any given historical moment is a different question. This paper makes no prediction about it. It only establishes what the path is, and what it is not.

— ✦ —

Conclusion: The Permanent Condition

Homo sapiens has not evolved meaningfully in fifty thousand years. The mind in the twenty-first century is the same mind that crossed the land bridge into the Americas, that painted the caves at Lascaux, that sat around the fires at which the first governing classes learned to tell the first subjugation stories. It was built for a world of small groups, known individuals, observable track records, and genuine communities. It carries that design into a world of anonymous millions, institutional symbols, unverifiable credentials, and manufactured belonging.

The five cognitive components described in this paper — the authority heuristic, the coalition loyalty instinct, the sacred value lock, the narrative capture, and the sunk cost of formation — were adaptive in the world the mind was built for. In the world we actually inhabit, they are the permanent exploit. Not a weakness. Not a design flaw. Features that were precision instruments in the Pleistocene, operating in an environment that has been deliberately engineered to misfire them.

The governing class discovered this exploit not through any single act of genius but through millennia of accumulated practical learning. The priests who found that the creation myth produced more reliable labor than the garrison were learning, inductively, which features of human cognition could be activated in service of extraction. The nobles who found that the anointing ceremony produced more reliable loyalty than the salary were learning the same thing from a different angle. The founding class that found that constitutional legitimacy produced more reliable obedience than divine right were doing the same thing with different materials in a different century.

They were all exploiting the same mind. The mind that you, reading this, are carrying. The same mind that is reading these words and, in the process, having the named mechanisms activated and perhaps, in some measure, partially counteracted by the naming. Not eliminated. Not repaired. Partially counteracted, for this moment, by the deliberate engagement of the slow system with the products of the fast one.

That is as much as description can do. It is not nothing. But it is also not freedom. Freedom from the exploit is not available to a species that carries the Pleistocene mind into the civilizational environment. What is available is the construction of the kinds of institutions, the kinds of distributed power, the kinds of genuinely competitive information environments, that make the exploitation expensive enough to constrain.

The governing class knows what it is doing. It has always known what it is doing — not in the explicit, theorized form that this series attempts, but in the practical, accumulated, institutionally-transmitted form of the craft of rule. The craft has been refined over fifty centuries by people who were rewarded for getting it right and eliminated for getting it wrong.

The craft works because the exploit is real.

The exploit is real because the mind is ancient.

The mind is ancient because evolution is slow.

Evolution is slow because we are not the problem the exploit poses. The problem the exploit poses is a problem of institutional design. The question is not: how do we change human nature? That is not available. The question is: how do we build and defend the arrangements that put friction between the ancient mind and the people who have spent fifty centuries learning to exploit it?

The series has been building this argument from the case studies up. This paper is the foundation that was always underneath them.

— ✦ —

Principal Sources and Further Reading

On Dual Process Cognition

Kahneman, Daniel. Thinking, Fast and Slow. New York: Farrar, Straus and Giroux, 2011. The accessible synthesis of dual-process theory. The core distinction between

System One (fast, automatic, pre-reflective) and System Two (slow, deliberate, effortful) is the paper's primary cognitive framework.

Evans, Jonathan St. B. T. and Keith Stanovich. 'Dual-Process Theories of Higher Cognition: Advancing the Debate.' Perspectives on Psychological Science 8, no. 3 (2013): 223–241. The scholarly state-of-the-art on the two-process framework and its limitations.

Sloman, Steven A. 'The Empirical Case for Two Systems of Reasoning.' Psychological Bulletin 119, no. 1 (1996): 3–22. The influential earlier formulation distinguishing associative and rule-based reasoning.

On Evolutionary Mismatch

Li, Norman P., Mark van Vugt, and Stephen M. Colarelli. 'The Evolutionary Mismatch Hypothesis: Implications for Psychological Science.' Current Directions in Psychological Science 27, no. 1 (2018): 38–44. The systematic treatment of evolved cognitive mechanisms operating in modern environments for which they were not calibrated.

Giphart, Ronald and Mark van Vugt. Mismatch: How Our Stone Age Brains Deceive Us Every Day (And What We Can Do About It). London: Robinson, 2018. The accessible treatment of the mismatch thesis across multiple domains.

Van Vugt, Mark and Anjana Ahuja. Naturally Selected: The Evolutionary Science of Leadership. New York: HarperBusiness, 2011. On the specific mismatch between evolved leadership psychology and modern institutional authority structures.

On Sacred Values and Moral Psychology

Tetlock, Philip E. 'Thinking the Unthinkable: Sacred Values and Taboo Cognitions.' Trends in Cognitive Sciences 7, no. 7 (2003): 320–324. The systematic treatment of values that resist cost-benefit reasoning.

Atran, Scott and Joseph Henrich. 'The Evolution of Religion: How Cognitive By-Products, Adaptive Learning Heuristics, Ritual Displays, and Group Competition Generate Deep Commitments to Prosocial Religions.' Biological Theory 5 (2010): 1–13. On the evolutionary origins of sacred value mechanisms.

Haidt, Jonathan. 'The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment.' Psychological Review 108, no. 4 (2001): 814–834. The

foundational paper arguing that moral judgment is primarily intuitive and that reasoning is typically post-hoc rationalization.

On Social Cognition and Group Size

Dunbar, Robin I. M. 'Coevolution of Neocortical Size, Group Size and Language in Humans.' Behavioral and Brain Sciences 16, no. 4 (1993): 681–694. The original formulation of the social brain hypothesis and the ~150 figure for human cognitive group size limits.

From the Patterns of Power Series

Bryan, Thomas J. 'Belief as Infrastructure: How Ancient Elites Engineered Narrative Solutions to the Problem of Scaled Extraction.' Patterns of Power, 2025. The nine-civilization comparative study that this paper provides the cognitive foundation for.

Bryan, Thomas J. 'The Durable Hierarchy.' Patterns of Power, 2026. First systematic treatment of the verification override.

Bryan, Thomas J. 'The Three-Layer Lock.' Patterns of Power, 2026. The Milgram experimental evidence and the Browning/Józefów case for the coalition loyalty and authority heuristic mechanisms.

Bryan, Thomas J. 'The Position Commands: Belief as the Primary Instrument of Governing-Class Power.' Patterns of Power, 2026. The companion paper on rational-legal authority architecture and the American constitutional case.

Bryan, Thomas J. 'Where Evil Lives: The Architect, the Soldier, and the Moral Architecture of War.' Patterns of Power, 2026. The moral implications of the gap between what architects know and what soldiers believe.

Bryan, Thomas J. 'Beyond the Accident.' Patterns of Power, 2026. The synthesis paper on institutional friction as the available response to the permanent condition this paper describes.

© 2026 Thomas J. Bryan
Creative Commons Attribution-NonCommercial 4.0 International License.
Patterns of Power: Systematic Analysis of Elite Control Mechanisms Across Civilizations.
Research and composition by Claude/Anthropic.